SIAM J. OPTIM.
Vol. 15, No. 4, pp. 1105–1127

A PREDICTOR-CORRECTOR ALGORITHM FOR LINEAR OPTIMIZATION BASED ON A SPECIFIC SELF-REGULAR PROXIMITY FUNCTION*

JIMING PENG†, TAMÁS TERLAKY†, AND YUNBIN ZHAO‡

Abstract. It is well known that the so-called first-order predictor-corrector methods working in a large neighborhood of the central path are among the most efficient interior-point methods (IPMs) for linear optimization (LO) problems. However, the best known iteration complexity of this type of method is $O(n \log \frac{(x^0)^T s^0}{\varepsilon})$. It is of interest to investigate whether the complexity of first-order predictor-corrector type methods can be further improved. In this paper, based on a specific self-regular proximity function, we define a new large neighborhood of the central path. In particular, we show that the new neighborhood matches the standard large neighborhood that is defined by the infinity norm and widely used in the IPM literature. A new first-order predictor-corrector method for LO that uses a search direction induced by self-regularity in corrector steps is proposed. We prove that our predictor-corrector algorithm, working in a large neighborhood, has an $O(\sqrt{n} \log n \log \frac{(x^0)^T s^0}{\varepsilon})$ iteration bound. Local superlinear convergence of the algorithm is also established.

Key words. linear optimization, interior-point methods, predictor-corrector method, self-regular proximity function, large neighborhoods, polynomial complexity, superlinear convergence

AMS subject classification. 90C05

DOI. 10.1137/040603991

1. Introduction.

To begin, we first clarify the notation used in this paper. Let $\Re^n$ denote the n-dimensional Euclidean space and e denote the vector with all components equal to 1. Further, $x \geq 0$ ($x > 0$) means that all components of x are nonnegative (positive). For any $x > 0$, $s > 0$, and real number r, the symbols xs, $\sqrt{x}$, and x^r denote vectors whose components are $x_i s_i$, $\sqrt{x_i}$, and x_i^r $(i = 1, \ldots, n)$, respectively. In particular, we have $x^{-1} = (\frac{1}{x_1}, \ldots, \frac{1}{x_n})^T$.

Since Karmarkar's seminal paper [8], a large body of research has been done in the field of IPMs for LO problems. Several recent books on this subject [19, 26, 27] contain substantial materials about various aspects of interior-point methods (IPMs) for linear optimization (LO). In the present paper, we consider the standard LO problem that takes the following form:

$$\min\{c^T x : Ax = b,\ x \geq 0\}.$$

*Received by the editors February 9, 2004; accepted for publication (in revised form) November 22, 2004; published electronically July 26, 2005. This research was supported by the MITACS project "New Interior Point Methods and Software for Convex Conic-Linear Optimization and Their Application to Solve VLSI Circuit Layout Problems" and an FPP grant from IBM Watson Research Laboratory.

http://www.siam.org/journals/siopt/15-4/60399.html

†Department of Computing and Software, McMaster University, Hamilton, Ontario, Canada, L8S 4L7 (pengj@mcmaster.ca, terlaky@mcmaster.ca). The research of the first author was partially supported by grant RPG 249635-02 of the National Sciences and Engineering Research Council of Canada (NSERC) and a PREA award. The research of the second author was supported by NSERC grant RPG 0048923 and the Canada Research Chair program.

‡Institute of Applied Mathematics, AMSS, Chinese Academy of Science, Beijing, China (ybzhao@amss.ac.cn). Part of the work was done when this author visited The Fields Institute for Research in Mathematical Sciences as a post doctor fellow. The research of this author was partially supported by grant 10201032 and grant 70221001 of the Natural Science Foundation of China.

Here $x \in \Re^n$, $c \in \Re^n$, $b \in \Re^m$, and A is an $m \times n$ matrix satisfying $\text{Rank}(A) = m$. Throughout the paper, we also assume [19] that there exists a strictly feasible primal-dual pair (x^0, y^0, s^0) such that

$$Ax^0 = b, \quad A^T y^0 + s^0 = c, \quad (x^0, s^0) > 0.$$

In the present paper, we focus on so-called first-order predictor-corrector IPMs that work in a large neighborhood of the central path.

We start with a brief review of predictor-corrector type methods. The idea of predictor-corrector type algorithms is very natural and elegant. These algorithms follow the central path by alternatively taking predictor steps and corrector steps. A predictor step aims at reducing the duality gap as much as possible, thus the resulting iterate might move close to the boundary of the feasible set, deviating from the central path, while a corrector step tries to bring the iterate back to a certain neighborhood of the central path. The best known representative of this class of methods is the Mizuno–Todd–Ye (MTY) algorithm for LO [11], which works in a small neighborhood of the central path. Quadratic convergence of the duality gap was proved in [29], and convergence of the iterates generated by the algorithm was proved in [4]. The MTY method was later extended to complementarity problems by Ye and Anstreicher [28] and to semidefinite optimization by Luo, Sturm, and Zhang [10]. According to numerical experiments, IPMs working in a large neighborhood of the central path perform much better than their counterparts in small neighborhoods [1]. Several authors have investigated IPMs that work in certain large neighborhoods of the central path [2, 3, 5, 9, 15, 16, 17, 24, 25, 31]. It should be noted that the so-far best known worst-case complexity for the first-order predictor-corrector algorithms with large neighborhoods is $O(n \log \frac{(x^0)^T s^0}{\epsilon})$, which is higher than that of those small neighborhood based algorithms. As remarked by Renegar [18], this becomes one of the ironies of the interior-point literature that algorithms which are more efficient in practice often have somewhat-worse complexity bounds.

Several strategies have already been proposed for improving the theoretical complexity of IPMs in large neighborhoods. For example, by using higher-order approximation of the central path, the iteration complexity of IPMs can become arbitrarily close to $O(\sqrt{n} \log \frac{(x^0)^T s^0}{n})$ [5, 7, 12, 17, 19, 20, 21, 30, 31]. In IPMs based on high-order approximation, we usually need to solve multiple linear systems and thus the computational cost per iteration increases as the order of the method increases. Recently, Peng, Roos, and Terlaky [13] proposed a class of IPMs in large neighborhoods and proved that some members of that family of IPMs have an $O(\sqrt{n} \log n \log \frac{(x^0)^T s^0}{\varepsilon})$ iteration bound. The approach in [13] is based on the notion of self-regular functions. Different from high-order approximation-type methods, IPMs based on self-regular functions solve only one linear system at each step. Inspired by the results in [13], we consider in the present paper the issue whether the iteration bound of first-order predictor-corrector type IPMs within a large neighborhood can be improved. This is far from being an easy task. As pointed out in [17], MTY-type predictor-corrector methods are much more difficult to develop and analyze in a wide neighborhood of the central path. The difficulty for developing MTY-type methods in large neighborhoods comes from two points. First, theoretically, corrector steps based on the classical first-order centering direction (first-order correctors) are rather inefficient in a large neighborhood.

A remedy for this point is to employ a corrector search direction based on self-

regular proximity functions introduced in [13]. However, this gives rise to another issue, namely whether the neighborhood defined by a self-regular proximity function matches the large neighborhoods used in the context of predictor-corrector methods. Second, once we use a new search direction in the corrector step, it is not clear if we can still obtain superlinear or quadratic convergence for the new algorithm.

The purpose of this paper is to address the above-mentioned questions and present a new first-order predictor-corrector method that works in a large neighborhood and has an $O(\sqrt{n}\log n\log\frac{(x^0)^Ts^0}{\varepsilon})$ iteration bound. To this end, a proximity measure function, induced by a specific self-regular function introduced in [13], will be explored. We then use this proximity function to define a new neighborhood of the central path and show that this new neighborhood matches the neighborhood $\mathcal{N}_\infty^-$ that has been widely used in most practical implementations of IPMs. Correspondingly, we also use the self-regularity induced search direction in the corrector steps. It is worth mentioning that we use the usual primal-dual affine scaling direction in the predictor step.

We note that in [23], the authors proposed a family of so-called generic central region IPMs that work in a large neighborhood of the central path and proved that the proposed algorithms enjoy an $O(\sqrt{n}\log\frac{(x^0)^Ts^0}{\epsilon})$ complexity. This approach was later extended to semidefinite optimization in [22], where a predictor-corrector version of the algorithm was also discussed. It should be noted that the predictor-corrector method of [22, 23] uses the second-order correction in the corrector step and thus it does not follow the same procedure as our algorithm.

The algorithm presented in this paper follows the prototype of MTY IPMs in a large neighborhood. In the predictor step, we take a step along the affine scaling search direction to reduce the duality gap, while in the corrector step we try to bring the iterate back to a certain neighborhood (might be a small neighborhood) of the central path.

This paper is organized as follows. In section 2, we introduce the new neighborhood of the central path induced by a self-regular proximity function and discuss some useful properties of the proximity function. In section 3, we describe our new algorithm and establish its $O(\sqrt{n}\log n\log\frac{(x^0)^Ts^0}{\varepsilon})$ complexity. Asymptotic superlinear convergence of the algorithm is proved in section 4. Finally, we close the paper by a few concluding remarks.

2. Proximity functions and neighborhoods. This section consists of two parts. In the first subsection, we discuss a specific self-regular proximity function and show that the neighborhood induced by this proximity function matches the large neighborhood defined by infinity norm and used in many practical implementations of IPMs. In the second subsection, we explore some properties of the proximity function.

2.1. Large neighborhoods and proximity functions. We note that under the interior-point assumption, for every given parameter $\mu > 0$ the system

$$\begin{aligned} Ax &= b, \quad x > 0, \\ A^Ty + s &= c, \quad s > 0, \\ xs &= \mu e, \end{aligned} \tag{2.1}$$

has a unique solution denoted by $(x(\mu), y(\mu), s(\mu))$. The union of solutions $\{x(\mu) : \mu > 0\}$ forms the so-called primal central path, while $\{(y(\mu), s(\mu)) : \mu > 0\}$ is called

the central path of the dual problem in the dual space. The iterates, generated by IPMs, are confined to be in certain neighborhoods of the central path, which are defined by certain proximity functions.

To describe these proximity functions, we need to introduce some notation. For any strictly feasible primal-dual pair (x, s) and any positive number μ,[1] we define

$$v := \sqrt{\frac{xs}{\mu}} \quad \text{and} \quad v^{-1} := \sqrt{\frac{\mu e}{xs}} \tag{2.2}$$

to be the vectors whose ith components are $\sqrt{\frac{x_i s_i}{\mu}}$ and $\sqrt{\frac{\mu}{x_i s_i}}$, respectively. Let us denote the parameter μ associated with the present duality gap by

$$\mu(x, s) = \frac{x^T s}{n}.$$

A neighborhood of the central path is usually defined by

$$\mathcal{N} := \{(x, s) > 0 : Ax = b,\ A^T y + s = c,\ \Phi(x, s, \mu(x, s)) \le \eta(\tau, n)\}, \tag{2.3}$$

where $\eta(\tau, n)$ is a function of a parameter τ and the dimension n of the underlying problem, and $\Phi(x, s, \mu(x, s))$ is some measurement to measure the distance from the current iterate to the targeted center.

There are many choices for the measurement function $\Phi(x, s, \mu)$ and the function $\eta(\tau, n)$ in (2.3). For instance, if we choose $\Phi(x, s, \mu) = \|\frac{xs}{\mu(x,s)} - e\|$ and $\eta(\tau, n) = \rho$ with $\rho \in (0, 1)$ (e.g., $\eta(\tau, n) = 0.25$), then we obtain a so-called small neighborhood. If we choose $\Phi(x, s, \mu) = \|\frac{xs}{\mu(x,s)} - e\|_\infty$ (or $\Phi(x, s, \mu) = \|(\frac{xs}{\mu(x,s)} - e)^-\|_\infty$ with $w^- = \min(w, 0)$) and $\eta(\tau, n) = \rho$ with $\rho \in (0, 1)$, then we get a so-called large neighborhood. Let us denote the neighborhoods generated via using the proximity functions $\|\frac{xs}{\mu(x,s)} - e\|_\infty$ and $\|(\frac{xs}{\mu(x,s)} - e)^-\|_\infty$ by $\mathcal{N}_\infty(\rho)$ and $\mathcal{N}_\infty^-(\rho)$, respectively. It is worth pointing out that the neighborhood $\mathcal{N}_\infty^-(\rho)$ is frequently used in practical implementations of IPMs, although the corresponding proximity function does not have a barrier property. On the other hand, as observed by many researchers, various IPMs are tightly associated with some proximity functions that have some barrier and coercive properties to prevent the iterate from moving to infinity and the boundary of the feasible set. For instance, Peng, Roos, and Terlaky [13] introduced the so-called self-regular proximity functions, which takes the form

$$\Phi(x, s, \mu) := \Psi(v) = \sum_{i=1}^{n} \psi(v_i), \tag{2.4}$$

where $\psi(t)$ is a so-called self-regular kernel function. For self-regular proximity functions, if we still choose $\eta(\tau, n)$ as a small constant, then the region defined by (2.3) becomes a small neighborhood. If $\eta(\tau, n) = \tau n = \mathcal{O}(n)$, then we obtain a large neighborhood.

However, as observed in [14], it is possible that the neighborhood $\mathcal{N}$ defined by (2.3) with some self-regular proximity function is too large and significantly differ-

[1] We alert the reader that throughout this paper μ is cast as an independent parameter, while the duality gap is associated with $\mu(x, s)$.

ent from the neighborhood $\mathcal{N}_\infty^-(\rho)$. To select an appropriate kernel function such that the resulting neighborhood defined by (2.3) can really match $\mathcal{N}_\infty^-(\rho)$, we consider the self-regular kernel function

$$\psi(t) = \frac{t^2-1}{2} + \frac{t^{-\log n}-1}{\log n}, \tag{2.5}$$

which is a self-regular function with $\nu_1 = 1$, $\nu_2 = \log n + 1$, $p = 1$, and $q = 1 + \log n$. The corresponding proximity function becomes

$$\Phi(x,s,\mu) = \Psi(v) = \frac{1}{2}\|v\|^2 - \frac{n}{2} + \sum_{i=1}^{n} \frac{v_i^{-\log n}-1}{\log n}. \tag{2.6}$$

One particular reason why we consider this specific proximity function is that, as shown in [13], among various large-update IPMs based on self-regular proximity functions, the algorithm with $q = O(\log n)$ has the best worst-case complexity result. Since $q = 1+\log n$ and the case for small q has been well studied in [13], in this paper we focus only on the case where q is sufficiently large, e.g., $q \geq 3$, which implies $n \geq 8$. Therefore, throughout this paper we assume $n \geq 8$.

Because the proximity function $\Psi(v)$ has an extremely strong barrier property, we need to choose correspondingly an appropriate function $\eta(\tau, n)$ in (2.3) so that the neighborhood given by (2.3) can match the neighborhood $\mathcal{N}_\infty^-(\rho)$ with $\rho \in (0,1)$. Let us define

$$\eta(\tau, n) := \frac{n^\tau - n}{\log n}. \tag{2.7}$$

Note that if $(x,s) \in \mathcal{N}_\infty^-(\rho)$ and $\mu = \mu(x,s)$, then we have $v_i^2 \geq 1-\rho$, which implies $-\log v_i \leq -\frac{1}{2}\log(1-\rho)$ for every $i \in \{1,\ldots,n\}$. From (2.6) we immediately obtain

$$\Phi(x,s,\mu(x,s)) = \Psi(v) = \sum_{i=1}^{n} \frac{v_i^{-\log n}-1}{\log n} = \sum_{i=1}^{n} \frac{n^{-\log v_i}-1}{\log n} \leq \frac{n^{1-\frac{1}{2}\log(1-\rho)}-n}{\log n}.$$

Our above discussion indicates that if we choose $\tau = 1 - \frac{1}{2}\log(1-\rho)$ in (2.7), then the neighborhood $\mathcal{N}$ defined by (2.3) contains the neighborhood $\mathcal{N}_\infty^-(\rho)$. It is worth mentioning that $\eta(\tau, n)$ is an increasing function of τ. In the rest of this paper, $\eta(\tau, n)$ is defined by (2.7) for some $\tau > 1$ and thus our algorithm will operate in a large neighborhood.

2.2. Properties of the proximity function. In this section we investigate various properties of the proximity function $\Phi(x,s,\mu)$. We first consider the case that $\mu = \mu(x,s)$. Recall that the value of $\Phi(x,s,\mu(x,s))$ is essentially determined by the sum $\sum_{i=1}^n v_i^{-\log n}$. To estimate this sum, we first observe that

$$u \geq 2 \Longrightarrow \frac{|t|^u - 1}{u} \geq \frac{t^2-1}{2} \quad \forall t \in \Re.$$

Since $\log n > 2$ if $n \geq 8$, we have

$$t^{-\log n} = \left(t^{-1}\right)^{\log n} \geq \frac{\log n}{2}(t^{-2}-1) + 1 \quad \forall t > 0,\ n \geq 8.$$

This relation and the fact that $\|v\|^2 = n$ when $\mu = \mu(x,s)$ imply[2]

$$\sum_{i=1}^{n} v_i^{-\log n} \geq \frac{\log n}{2}(\|v^{-1}\|^2 - n) + n = \frac{\log n}{2}\|v - v^{-1}\|^2 + n \geq n \quad \forall n \geq 8, \tag{2.8}$$

and equality holds if and only if $v = e$.

We proceed to explore more properties of the proximity function $\Phi(x,s,\mu)$. For fixed $(x,s) > 0$, there exists $\tau \geq 1$ such that

$$\Phi(x,s,\mu(x,s)) = \sum_{i=1}^{n} \frac{v_i^{-\log n} - 1}{\log n} = \eta(\tau, n),$$

which equals

$$\tau = \frac{\log(\sum_{i=1}^{n} v_i^{-\log n})}{\log n}.$$

The above relation indicates that τ provides another way to measure the proximity function. Let us define

$$f(\theta) := \Phi(x,s,\theta\mu(x,s)).$$

It is easy to see that $f(\theta)$ is convex for $\theta > 0$. Denote

$$\theta^* = \arg\min_{\theta>0} f(\theta).$$

It follows from the optimality condition of $f(\theta)$ that

$$\frac{n}{2(\theta^*)^2} - \frac{(\theta^*)^{\frac{1}{2}\log n}}{2\theta^*} \sum_{i=1}^{n} v_i^{-\log n} = 0. \tag{2.9}$$

From this equality and (2.7) we obtain

$$n^{1-\tau} = (\theta^*)^{1+\frac{1}{2}\log n},$$

which yields

$$(1-\tau)\log n = \left(1 + \frac{1}{2}\log n\right)\log \theta^*.$$

Thus we have

$$\theta^* = \exp^{-\frac{(2\tau-2)\log n}{2+\log n}}. \tag{2.10}$$

For a given primal-dual pair (x,s), we next consider the behavior of the proximity function $\Phi(x,s,\mu)$ as a function of μ. First, we mention that it is trivial to

[2]One can easily derive that $\Phi(x,s,\mu(x,s)) \geq \frac{1}{2}\|v - v^{-1}\|^2$, and the right-hand side is a well-known proximity measure in the literature, which is also a self-regular proximity with $p = 1$ and $q = 3$; see, e.g., [13].

verify that Φ is convex in μ. From the choice (2.10) of θ^*, we immediately have the following lemma.

LEMMA 2.1. *Given a primal-dual pair (x, s), suppose that $\Phi(x, s, \mu(x, s)) = \eta(\tau, n)$ for some $\tau > 1$ where $\eta(\tau, n)$ is given by (2.7), and θ^* is given by (2.10). Then the proximity function $\Phi(x, s, \mu)$ is convex with respect to μ and attains its global minimum at*

$$\mu^* = \theta^* \mu(x, s).$$

In particular, we have

$$\sum_{i=1}^{n} \left(\frac{x_i s_i}{\mu} \right)^{-\frac{1}{2}\log n} - \frac{1}{\mu} \sum_{i=1}^{n} x_i s_i \leq 0, \iff 0 < \mu \leq \mu^*,$$

and the equality holds if and only if $\mu = \mu^*$.

By using (2.10) one can easily show that

$$\exp^{2-2\tau} \leq \theta^*. \tag{2.11}$$

Further, we have

$$\theta^* \leq \exp^{1-\tau} \quad \forall n \geq 8. \tag{2.12}$$

The above two bounds indicate that if $n \geq 8$, θ^* is uniformly bounded above and below and these bounds depend only on τ. Let us cast θ^* as a function of τ and n, i.e.,

$$\theta^* := \theta(\tau, n) = \exp^{-\frac{(2\tau-2)\log n}{2+\log n}}. \tag{2.13}$$

We point out that we use $\mu = \theta^* \mu(x, s)$ in the update of the parameter μ in our algorithm. Lemma 2.1 indicates that if $\Phi(x, s, \mu(x, s)) = \eta(\tau, n)$ with $n \geq 8$ and we use $\mu = \theta(\tau, n)\mu(x, s)$ (which is indeed a large update) as our targeted μ in the algorithm, then the proximity function $\Phi(x, s, \mu)$ will not increase after the update. It is straightforward to verify that $\theta(\tau, n)$ is a decreasing function of τ.

For fixed $(x, s) > 0$ and $\theta(\tau, n)$ given by (2.13), let us define

$$\eta_1(\tau, n) := \left(\frac{n}{2} + \frac{n}{\log n} \right) \left(\frac{1}{\theta(\tau, n)} - 1 \right). \tag{2.14}$$

If $\Phi(x, s, \mu(x, s)) = \eta(\tau, n)$, then from (2.9) we can conclude

$$\Phi(x, s, \theta(\tau, n)\mu(x, s)) = \eta_1(\tau, n).$$

It is trivial to verify that $\eta_1(\tau, n)$ is strictly increasing with respect to τ. This gives the following corollary.

COROLLARY 2.2. *Let $(x, s) > 0$, $n \geq 8$, and $\tau \geq 2$, and suppose that the proximity function $\Phi(x, s, \mu)$ attains its minimum at μ^*. Then $\Phi(x, s, \mu(x, s)) \leq \eta(\tau, n)$ if and only if $\Phi(x, s, \mu^*) \leq \eta_1(\tau, n)$.*

We also point out here that when $n \geq 8$, $\frac{\eta_1(n,\tau)}{n}$ is uniformly bounded below and above by some constants independent of n.

It should be mentioned that in the above discussion, we sometimes implicitly assume that the parameter τ is dependent on the present iterate (x, s). However, in most IPMs, we usually fix the parameter in the definition of the neighborhood (such as ρ in $\mathcal{N}_\infty^-(\rho)$) and only require the iterate to be in a certain neighborhood. Therefore, in the remaining part of this paper, we only assume that $\tau \geq 2$ is fixed. If the present iterate is in a certain neighborhood (say $\Phi(x, s, \mu(x, s)) \leq \eta(\tau, n)$), then we update the parameter μ and compute a search direction w.r.t. the targeted μ. In this case, we need to investigate the change of the proximity function. Our following result provides an upper bound for $\Phi(x, s, \theta(\tau, n)\mu(x, s))$ whenever $\Phi(x, s, \mu(x, s)) < \eta(\tau, n)$.

LEMMA 2.3. *If $\Phi(x, s, \mu(x, s)) \leq \eta(\tau, n)$, then one has*

$$\Phi(x, s, \theta(\tau, n)\mu(x, s)) \leq \eta_1(\tau, n).$$

Proof. From the assumption in the lemma, without loss of generality, we can assume

$$\Phi(x, s, \mu(x, s)) = \eta(\tau_0, n) \leq \eta(\tau, n)$$

for some $\tau_0 \leq \tau$. Since $\theta(\tau, n)$ is decreasing in τ, we therefore have $\theta(\tau_0, n) \geq \theta(\tau, n)$. Now recall the definition of $\theta(\tau_0, n)$. From the optimality condition (2.9) we have

$$\frac{x^T s}{\theta(\tau_0, n)\mu(x, s)} = \sum_{i=1}^{n} \left(\frac{x_i s_i}{\theta(\tau_0, n)\mu(x, s)} \right)^{-\frac{1}{2}\log n}.$$

It follows that

$$\begin{aligned} \sum_{i=1}^{n} \left(\frac{x_i s_i}{\theta(\tau, n)\mu(x, s)} \right)^{-\frac{1}{2}\log n} &= \left(\frac{\theta(\tau, n)}{\theta(\tau_0, n)} \right)^{\frac{1}{2}\log n} \sum_{i=1}^{n} \left(\frac{x_i s_i}{\theta(\tau_0, n)\mu(x, s)} \right)^{-\frac{1}{2}\log n} \\ &\leq \sum_{i=1}^{n} \left(\frac{x_i s_i}{\theta(\tau_0, n)\mu(x, s)} \right)^{-\frac{1}{2}\log n} = \frac{x^T s}{\theta(\tau_0, n)\mu(x, s)} \\ &\leq \frac{x^T s}{\theta(\tau, n)\mu(x, s)} = \frac{n}{\theta(\tau, n)}. \end{aligned} \tag{2.15}$$

Applying this inequality to the proximity function $\Phi(x, s, \mu)$, we have

$$\begin{aligned} \Phi(x, s, \theta(\tau, n)\mu(x, s)) &= \frac{n}{2}\left(\frac{1}{\theta(\tau, n)} - 1 \right) + \frac{1}{\log n}\left(\sum_{i=1}^{n} \left(\frac{x_i s_i}{\theta(\tau, n)\mu(x, s)} \right)^{-\frac{1}{2}\log n} - n \right) \\ &\leq \frac{n}{2}\left(\frac{1}{\theta(\tau, n)} - 1 \right) + \frac{1}{\log n}\left(\frac{n}{\theta(\tau, n)} - n \right) = \eta_1(\tau, n), \end{aligned}$$

which finishes the proof of the lemma. □

We present another lemma that characterizes the behavior of the proximity function $\Phi(x, s, \mu)$.

LEMMA 2.4. *Let* $(x,s) > 0$, $n \geq 8$, *and* $\tau \geq 2$. *Then the relation* $\Phi(x,s,\mu(x,s)) \leq \eta(\tau,n)$ *holds if and only if there exists a parameter* $\mu > 0$ *such that* $\Phi(x,s,\mu) \leq \eta_1(\tau,n)$.

Proof. The necessary part follows directly from Lemma 2.3, thus it remains to consider the sufficient part. Suppose that the proximity function $\Phi(x,s,\mu)$ has a global minimum at μ^*. The assumption in the lemma implies

$$\Phi(x,s,\mu^*) \leq \eta_1(\tau,n).$$

It follows immediately from Corollary 2.2

$$\Phi(x,s,\mu(x,s)) \leq \eta(\tau,n).$$

This completes the proof of the lemma. ∎

We close this section with the following corollary, which gives an estimation of μ^* whenever $\Phi(x,s,\mu(x,s) \leq \eta(\tau,n))$. The corollary follows from Lemma 2.4 and the fact that $\theta(\tau,n)$ is decreasing with respect to τ.

COROLLARY 2.5. *Let* $(x,s) > 0$, $n \geq 8$, *and* $\tau \geq 2$, *and suppose that the proximity function* $\Phi(x,s,\mu)$ *attains its minimum at* μ^*. *Then* $\Phi(x,s,\mu(x,s)) \leq \eta(\tau,n)$ *if and only if*

$$\mu^* \geq \theta(\tau,n)\mu(x,s).$$

3. A pseudo predictor-corrector method. In this section, we describe our algorithm and establish its complexity result. The section is divided into three parts. In the first subsection, we describe our new algorithm. The second subsection is devoted to investigating the behavior of the proximity function in the corrector step. In the last part, we estimate the step size used in the predictor step and summarize the complexity result.

3.1. The algorithm. We start with a brief description of standard IPMs. Suppose that the present iterate (x,s) is strictly feasible. For any given duality gap parameter μ, we can apply Newton method to system (2.1) and obtain a search direction via solving the following linear system of equations:

$$\begin{aligned} A\Delta x &= 0, \\ A^T\Delta y + \Delta s &= 0, \\ x\Delta s + s\Delta x &= \mu e - xs. \end{aligned} \tag{3.1}$$

Let us consider the case that v is defined by (2.2) with $\mu = \theta(\tau,n)\mu(x,s)$. Therefore the relation $\Phi(x,s,\theta(\tau,n)\mu(x,s)) = \Psi(v)$ holds. We denote the search direction in the scaled v-space as

$$d_x := \frac{v\Delta x}{x} \quad \text{and} \quad d_s := \frac{v\Delta s}{s}. \tag{3.2}$$

Using this notation and (2.2), we can write system (3.1) in the scaled v-space as

$$\begin{aligned} \bar{A}d_x &= 0, \\ \bar{A}^T\Delta y + d_s &= 0, \\ d_x + d_s &= v^{-1} - v, \end{aligned} \tag{3.3}$$

where $\bar{A} = \frac{1}{\mu}AV^{-1}X$, $V = \text{diag}\,(v)$, and $X = \text{diag}\,(x)$. Recall that the so-called primal-dual affine scaling search direction is the solution of the system

$$\begin{aligned} \bar{A}d_x &= 0, \\ \bar{A}^T\Delta y + d_s &= 0, \\ d_x + d_s &= -v, \end{aligned} \tag{3.4}$$

while the search direction induced by the self-regular proximity function $\Psi(v)$ can be computed by solving [13]

$$\begin{aligned} \bar{A}d_x &= 0, \\ \bar{A}^T\Delta y + d_s &= 0, \\ d_x + d_s &= v^{-1-\log n} - v = -\nabla\Psi(v). \end{aligned} \tag{3.5}$$

The following technical result shows that if the present iterate is in the neighborhood $\mathcal{N}$ defined by (2.3), then the duality gap will not increase along the search direction obtained by solving (3.5).

PROPOSITION 3.1. *Let (x, s) be a strictly feasible pair and $\Psi(v) \le \eta_1(\tau, n)$, where $\mu = \theta(\tau, n)\mu(x, s)$ and $\eta_1(\tau, n)$ is defined by* (2.14). *Then the solution (d_x, d_s) of system* (3.5) *satisfies*

$$v^T(d_x + d_s) \le 0,$$

and the equality holds if and only if $\Psi(v) = \eta_1(\tau, n)$.

Proof. Since $\Psi(v) \le \eta_1(\tau, n)$, from Lemma 2.4 we obtain

$$\Phi(x, s, \mu(x, s)) \le \eta(n, \tau).$$

Let μ^* be the global minimizer of $\Phi(x, s, \mu)$ with respect to μ. From Corollary 2.5 we conclude

$$\mu^* \ge \theta(\tau, n)\mu(x, s).$$

The proposition follows immediately from the last conclusion of Lemma 2.1 and (3.5). ∎

The above proposition indicates that if we take a step along the search direction obtained from (3.5) with $\mu = \mu^*$, then the duality gap remains constant and it decreases if and only if the target value $\mu \le \mu^*$.

Now we are ready to describe our algorithm. To start the algorithm, we assume that a strictly feasible starting point is available and the starting point is in the neighborhood $\mathcal{N}$ where $\eta(\tau, n)$ is defined by (2.7). Then we use a corrector-type step to reduce the value of the proximity function and thus bring the iterate into a smaller neighborhood of the central path. We would like to alert the reader to the distinction between our corrector step and the standard corrector step in the IPM literature. In a standard corrector step, the search direction is always computed by solving (3.1) with $\mu = \mu(x, s)$ and thus the duality gap remains invariant. However, as stated in Proposition 3.1, the duality gap in a corrector step of Algorithm 3.1 will never increase but might decrease. Each corrector step is followed by a predictor step that aims at reducing the duality gap while keeping the iterate in the large neighborhood $\mathcal{N}$. This process is repeated until the duality gap is small enough.

The procedure of our algorithm is outlined as follows. In the procedure, we use the notation $x(\alpha) = x + \alpha\Delta x$, $y(\alpha) = y + \alpha\Delta y$, and $s(\alpha) = s + \alpha\Delta s$.

Algorithm 3.1

Input:
Proximity parameters $\tau \geq 2$, $\theta(\tau, n)$ given by (2.13),
and $\eta(\tau, n)$ given by (2.7);
an accuracy parameter $\varepsilon > 0$;
$(x, s) = (x^0, s^0)$ such that $\Phi(x, s, \mu(x, s)) \leq \eta(\tau, n)$.

begin
while $x^T s \geq \varepsilon$ **do**
begin
corrector step
Compute $\mu(x, s)$ and update μ to $\mu := \theta(\tau, n)\mu(x, s)$;
Solve the system (3.5) for $(\Delta x, \Delta y, \Delta s)$;
Find a feasible step size α that reduces
the proximity function sufficiently;[a]
Update the iterate by $x = x(\alpha)$, $y = y(\alpha)$, $s = s(\alpha)$;
Predictor step
Solve the system (3.4) for $(\Delta x, \Delta y, \Delta s)$,
Find a sufficiently large step size $\alpha \in (0, 1]$
such that[b] $\Phi(x(\alpha), s(\alpha), \mu(x, s)(\alpha)) \leq \eta(\tau, n)$;
Update the iterate by $x = x(\alpha)$, $y = y(\alpha)$, $s = s(\alpha)$;
end
end

[a] In Theorem 3.5, we shall specify a step size α that can reduce the proximity function sufficiently.

[b] In Theorem 3.7, we shall specify a step size α that keeps the iterate in the neighborhood.

3.2. The corrector step. In this subsection we estimate the change of the proximity function in the corrector step. For this we first discuss how to choose a suitable step size to reduce the proximity function Φ in the corrector step. Suppose that the current point (x, s) is in the neighborhood $\mathcal{N}$ (given by (2.3)) where $\eta(\tau, n)$ is given by (2.7), and let

$$\sigma = \|\nabla \Psi(v)\| .$$

Note that when $\log n \geq 1$, we have

$$\sigma \geq \left\| v - v^{-1} \right\| . \tag{3.6}$$

In what follows we estimate the magnitudes of σ and $v_{\min}$. In particular, we will derive a lower bound of $v_{\min}$, from which we can further get an estimate of the maximal feasible step size.

LEMMA 3.2. *Suppose that the present iterate (x, s) is in the neighborhood $\mathcal{N}$ defined by (2.3) and v is defined by (2.2) with $\mu = \theta(\tau, n)\mu(x, s)$. If $n \geq 8$ and $\tau \geq 2$, then one has*

$$v_{\min} \geq \exp^{-\frac{(2\tau - 2)\log n}{2 + \log n}} \geq \exp^{-\tau}, \tag{3.7}$$

$$\sigma \geq \frac{5}{6} v_{\min}^{-1-\log n}. \tag{3.8}$$

Proof. Using the notation v, we can write $\Phi(x,s,\theta(\tau,n)\mu(x,s)) = \Psi(v)$. When $(x,s) \in \mathcal{N}$, from Lemma 2.3 we obtain $\Psi(v) \le \eta_1(\tau,n)$, where $\eta_1(\tau,n)$ is given by (2.14). It follows from (2.15) that

$$v_{\min}^{-\log n} < \sum_{i=1}^{n} v_i^{-\log n} \le \sum_{i=1}^{n} (v_i)^2 = \frac{x^T s}{\theta(\tau,n)\mu(x,s)} = n \exp^{\frac{(2\tau-2)\log n}{2+\log n}}.$$

Therefore, one has

$$\log v_{\min}^{-1} < 1 + \frac{2\tau - 2}{2 + \log n} \le \tau,$$

which further yields (3.7).

To prove (3.8), it suffices to consider the case where $v_{\min} \le 1$. Observe that if $n \ge 8$ and $\tau \ge 2$, then from (3.6) and (2.12) we obtain

$$\sigma^2 \ge \left\|v - v^{-1}\right\|^2 > \|v\|^2 - 2n = \left(\frac{1}{\theta(\tau,n)} - 2\right) n \ge \left(\exp^{\tau-1} - 2\right) n \ge 6 \quad \forall n \ge 8,\ \tau \ge 2.$$

This gives (3.8) when

$$v_{\min}^{1+\log n} > \frac{1}{6}.$$

If

$$v_{\min}^{1+\log n} \le \frac{1}{6},$$

then we have

$$\begin{aligned} v_{\min}^{1+\log n}\sigma &\ge v_{\min}^{1+\log n}(v_{\min}^{-1-\log n} - v_{\min}) \\ &= 1 - v_{\min}^{2+\log n} \ge 1 - v_{\min}^{1+\log n} \\ &\ge \frac{5}{6}. \end{aligned}$$

The proof of the lemma is finished. ∎

LEMMA 3.3. *Given $t, w \in [0,1)$. If $-\log(1-t) \ge w$, then one has $t \ge \frac{w}{1+w}$.*

Proof. Since $t \in [0,1)$, one has

$$w \le -\log(1-t) = \log\left(1 + \frac{t}{1-t}\right) \le \frac{t}{1-t},$$

which further gives the desired relation in the lemma. ∎

The following technical lemma from [13] will be used in our later analysis. For self-completeness, we include it here without proof.

LEMMA 3.4. *Suppose that $h(t)$ is a twice differentiable convex function with $h(0) = 0$, $h'(0) < 0$ and that $h(t)$ attains its global minimum at its stationary point $t^* > 0$. If $h''(t)$ is increasing with respect to t, then for any $t \in [0, t^*]$ we have*

$$h(t) \le \frac{h'(0)t}{2}.$$

We proceed to specify a step size that reduces the proximity function sufficiently in the corrector step of Algorithm 3.1. We mention that by using Theorem 1 in [13],

we can get a step size α that reduces the proximity function value. However, the estimation in [13] focuses on the relation of the reduction and the proximity function itself, while our following theorem tries to estimate the reduction of the proximity function by using the information of the new updated iterate. For self-completeness, we also present a detailed proof here.

THEOREM 3.5. *Suppose that the present iterate (x, s) is in the neighborhood $\mathcal{N}$ defined by (2.3) and v is defined by (2.2) with $\mu = \theta(\tau, n)\mu(x, s)$. Let (d_x, d_s) be the solution of system (3.5) with $n \geq 8$ and $\tau \geq 2$. Then the step size $\alpha = \frac{v_{\min}}{3\sigma \log n}$ is strictly feasible. Moreover, for this step size, we have*

$$\Phi(x(\alpha), s(\alpha), \mu) \leq \Phi(x(0), s(0), \mu) - \frac{5}{66 \log n} \max\{v_{\min}\sigma, v_{\min}^{-\log n}(\alpha)\}.$$

Proof. To prove the theorem, we need to estimate the maximal feasible step size $\alpha_{\max}$. By using the orthogonality of d_x and d_s, we have

$$\max\{\|d_x\|^2, \|d_s\|^2\} \leq \|d_x\|^2 + \|d_s\|^2 = \|d_x + d_s\|^2 = \sigma^2. \tag{3.9}$$

From (3.2) we conclude that the strict feasibility of $(x+\alpha\Delta x, s+\alpha\Delta s)$ can be retained if and only if $(v+\alpha d_x, v+\alpha d_s)$ is strictly positive. This is certainly true if $v_{\min}-\alpha\sigma > 0$. This gives

$$\alpha_{\max} \geq v_{\min}\sigma^{-1} \geq \sigma^{-1} \exp^{-\tau}.$$

We progress to investigate the behavior of the function $\Phi(x(\alpha), s(\alpha), \theta(\tau, n)\mu(x, s))$ along the search direction $(\Delta x, \Delta s)$ obtained by solving (3.5). Let us define

$$v(\alpha) = \sqrt{(v + \alpha d_x)(v + \alpha d_s)}.$$

Using this, we can write the proximity function $\Phi(x(\alpha), s(\alpha), \theta(\tau, n)\mu(x, s))$ in the scaled v-space as

$$\Phi(x(\alpha), s(\alpha), \theta(\tau, n)\mu(x, s)) = \Psi(v(\alpha)).$$

Note that from the choice of $\Psi(v(\alpha))$, we have

$$\begin{aligned}
\Psi(v(\alpha)) &= \frac{(v + \alpha d_x)^T(v + \alpha d_s)}{2} - \frac{n}{2} + \frac{1}{\log n}\sum_{i=1}^{n}\left(v_i(\alpha)^{-\log n} - 1\right) \\
&\leq \frac{1}{2}\|v\|^2 + \alpha v^T(d_x + d_s) - \frac{n}{2} \\
&\quad + \frac{1}{2\log n}\sum_{i=1}^{n}((v + \alpha d_x)_i^{-\log n} + (v + \alpha d_s)_i^{-\log n} - 2),
\end{aligned}$$

where the inequality follows from the orthogonality of d_x and d_s as well as the self-regularity of the function $\psi(t)$. Let us cast the difference between the proximity functions before and after a step as a function of the step size α; then from (3.5) we obtain

$$\begin{aligned}
g(\alpha) &:= \Psi(v(\alpha)) - \Psi(v(0)) \\
&\leq -\alpha\sum_{i=1}^{n}(v_i^2 - v_i^{-\log n}) \\
&\quad + \frac{1}{2\log n}\sum_{i=1}^{n}((v + \alpha d_x)_i^{-\log n} + (v + \alpha d_s)_i^{-\log n} - 2v_i^{-\log n}).
\end{aligned}$$

One can easily verify that $g(\alpha)$ is convex. Furthermore, by simple calculus one gets

$$g(0) = 0 \quad \text{and} \quad g'(0) = -\sigma^2.$$

For any $\alpha \le \alpha_{\max}$, we have

$$\begin{aligned} g''(\alpha) &= (1+\log n)\sum_{i=1}^{n}((d_x^i)^2(v+\alpha d_x)_i{}^{-2-\log n} + (d_s^i)^2(v+\alpha d_s)_i{}^{-2-\log n}) \\ &\le (1+\log n)\sum_{i=1}^{n}((d_x^i)^2(v_i-\alpha\sigma)^{-2-\log n} + (d_s^i)^2(v_i-\alpha\sigma)^{-2-\log n}) \\ &\le (1+\log n)\sigma^2(v_{\min}-\alpha\sigma)^{-2-\log n}, \end{aligned}$$

where the inequalities follow from the fact $\sigma = \|d_x + d_s\|$ and

$$\min_{i=1,\dots,n}\{v_i+\alpha d_x^i, v_i+\alpha d_s^i\} \ge \min_{i=1,\dots,n} v_i - \alpha\sigma \ge v_{\min} - \alpha\sigma.$$

Let

$$g_1(\alpha) = -\alpha\sigma^2 + \int_0^\alpha \int_0^\zeta (1+\log n)\sigma^2(v_{\min}-t\sigma)^{-2-\log n} dt d\zeta.$$

It is easy to see that $g_1(\alpha)$ is also convex in the feasible region with

$$g_1(0) = 0 \quad \text{and} \quad g_1'(0) = -\sigma^2.$$

Moreover, since $g_1''(\alpha) \ge g''(\alpha)$, one has

$$g(\alpha) \le g_1(\alpha).$$

In what follows we estimate the step size α^* at which $g_1(\alpha)$ has a global minimum. From the convexity of $g_1(\alpha)$ it follows that α^* must satisfy $g_1'(\alpha) = 0$, which gives

$$-\sigma^2 + \sigma((v_{\min}-\alpha^*\sigma)^{-1-\log n} - v_{\min}{}^{-1-\log n}) = 0, \tag{3.10}$$

or equivalently

$$-\sigma v_{\min}^{1+\log n} + ((1-\alpha^* v_{\min}^{-1}\sigma)^{-1-\log n} - 1) = 0.$$

Using (3.8), we derive

$$\frac{11}{6} \le \left(1-\alpha^* v_{\min}^{-1}\sigma\right)^{-1-\log n}, \tag{3.11}$$

which is equal to

$$-(1+\log n)\log(1-\alpha^* v_{\min}^{-1}\sigma) \ge \log\frac{11}{6} > \frac{3}{5}.$$

Since $\log n \ge 2$, the above relation gives

$$-\log(1-\alpha^* v_{\min}^{-1}\sigma) \ge \frac{3}{5+5\log n} \ge \frac{2}{5\log n}.$$

Using Lemma 3.3 and the fact that $\log n \geq 2$, we derive

$$\alpha^* v_{\min}^{-1}\sigma \geq \frac{2}{5\log n + 2} \geq \frac{1}{3\log n}.$$

Now let us focus on the case where the step size

$$\alpha = \frac{v_{\min}}{3\sigma \log n}. \tag{3.12}$$

By using Lemma 3.4, we have

$$\Psi(v(\alpha)) \leq \Psi(v) + g_1(\alpha) \leq \Psi(v) - \frac{\alpha\sigma^2}{2} \leq \Psi(v) - \frac{v_{\min}\sigma}{6\log n}. \tag{3.13}$$

On the other hand, because $\alpha \leq \alpha^*$, it must hold $g_1'(\alpha) \leq 0$, i.e.,

$$-\sigma^2 + \sigma((v_{\min} - \alpha\sigma)^{-1-\log n} - v_{\min}^{-1-\log n}) \leq 0. \tag{3.14}$$

It follows that

$$(v_{\min} - \alpha\sigma)^{-1-\log n} \leq \sigma + v_{\min}^{-1-\log n} \leq \frac{11}{5}\sigma,$$

where the last inequality is given by (3.8). From (3.13) we immediately obtain

$$\Psi(v(\alpha)) - \Psi(v) \leq -\frac{v_{\min}\sigma}{6\log n} \leq -\frac{5v_{\min}}{66\log n}(v_{\min} - \alpha\sigma)^{-1-\log n} \leq -\frac{5(v_{\min} - \alpha\sigma)^{-\log n}}{66\log n}.$$

Notice that for all feasible step size α, we have

$$v_{\min}(\alpha) = \min_{1\leq i\leq n} \sqrt{(v + \alpha d_x)(v + \alpha d_s)} \geq v_{\min} - \alpha\sigma.$$

It follows from the above two relations that

$$\Psi(v(\alpha)) \leq \Psi(v) - \frac{5v_{\min}^{-\log n}(\alpha)}{66\log n}.$$

Combining the above inequality with (3.13), we obtain the desired inequality in the theorem. □

3.3. The predictor step and complexity of the algorithm. In this subsection we estimate the step size used in the predictor step that keeps the new iterate in the neighborhood $\mathcal{N}$ defined by (2.3). To guarantee the resulting iterate in a certain neighborhood, it suffices to show that the value of the proximity function for a strictly feasible step size is bounded above by $\eta(\tau, n)$ or $\eta_1(\tau, n)$, depending on the duality parameter μ used in the definition of the proximity function. In order to distinguish the iterates before and after a corrector step, let (x^+, s^+) denote the primal-dual pair after the corrector step. It should be mentioned that in the predictor step, we have the freedom to choose the reference duality gap parameter μ because in the predictor step, the resulting search directions Δx and Δs in the original x- and s-spaces are completely independent of the parameter μ. Let μ^+ be the reference duality gap parameter used in the predictor step. Correspondingly, the scaled vector v^+ can be defined as

$$v^+ = \sqrt{\frac{x^+ s^+}{\mu^+}} \quad \text{and} \quad (v^+)^{-1} = \sqrt{\frac{\mu^+ e}{x^+ s^+}}. \tag{3.15}$$

In what follows we will discuss how to choose an appropriate duality gap parameter μ^+ such that the resulting scaled vector v^+ satisfies the following inequalities:

$$\|v^+\|^2 \le \frac{n}{\theta(\tau,n)}, \tag{3.16}$$

$$\sum_{i=1}^{n} (v_i^+)^{-\log n} - \|v^+\|^2 \le 0, \tag{3.17}$$

$$\Psi(v^+) \le \Phi(x,s,\theta(\tau,n)\mu(x,s)) - \frac{5}{66\log n}(v_{\min}^+)^{-\log n}. \tag{3.18}$$

The following lemma proves the existence of such a parameter μ^+.

LEMMA 3.6. *Let $n \ge 8$, $\tau \ge 2$, and (x^+, s^+) be the iterate after the corrector step and $\mu(x,s)^+ = (x^+)^T s^+/n$. Then there exists a duality gap parameter μ^+ such that the vector v^+ given by* (3.15) *satisfies inequalities* (3.16)–(3.18).

Proof. To prove the lemma, we first observe that if

$$\sum_{i=1}^{n} \left(\frac{x_i^+ s_i^+}{\theta(\tau,n)\mu(x,s)} \right)^{-\frac{1}{2}\log n} - \frac{(x^+)^T s^+}{\theta(\tau,n)\mu(x,s)} \le 0, \tag{3.19}$$

then we can set $\mu^+ = \theta(\tau,n)\mu(x,s)$, the same as in the corrector step. In this case, (3.19) gives (3.17) and (3.18) follows from Theorem 3.5. Moreover, as indicated by Proposition 3.1, the duality gap will not increase in the corrector step; this implies $\mu^+ \ge \theta(\tau,n)\mu(x,s)^+$, and thus inequality (3.16) holds as well.

Thus it remains to consider the case where

$$\sum_{i=1}^{n} \left(\frac{x_i^+ s_i^+}{\theta(\tau,n)\mu(x,s)} \right)^{-\frac{1}{2}\log n} - \frac{(x^+)^T s^+}{\theta(\tau,n)\mu(x,s)} > 0 \tag{3.20}$$

after the corrector step. Now let us cast $\Phi(x^+, s^+, \mu)$ as a function of μ. Since $\Phi(x^+, s^+, \mu)$ is a convex function of μ, it must attain its global minimum at some point μ^* satisfying

$$\sum_{i=1}^{n} \left(\frac{x_i^+ s_i^+}{\mu^*} \right)^{-\frac{1}{2}\log n} - \frac{(x^+)^T s^+}{\mu^*} = 0.$$

Because $\log n \ge 2$, (3.20) implies

$$\mu^* < \theta(\tau,n)\mu(x,s).$$

On the other hand, if we choose $\mu^+ = \mu^*$ and define v^+ by (3.15), then from Theorem 3.5 we obtain

$$\begin{aligned}
\Psi(v^+) &\le \Phi(x^+, s^+, \theta(\tau,n)\mu(x,s)) \\
&\le \Phi(x,s,\theta(\tau,n)\mu(x,s)) - \frac{5}{66\log n} \max_{i=1,\dots,n} \left\{ \left(\sqrt{\frac{x_i^+ s_i^+}{\theta(\tau,n)\mu(x,s)}} \right)^{-\log n} \right\} \\
&\le \Phi(x,s,\theta(\tau,n)\mu(x,s)) - \frac{5}{66\log n} \max_{i=1,\dots,n} \left\{ \left(\sqrt{\frac{x_i^+ s_i^+}{\mu^*}} \right)^{-\log n} \right\} \\
&= \Phi(x,s,\theta(\tau,n)\mu(x,s)) - \frac{5}{66\log n}(v_{\min}^+)^{-\log n}.
\end{aligned}$$

It remains to verify that (3.16) holds for this special choice. Note that because

$$\Phi(x^+, s^+, \mu^+) < \eta_1(\tau, n),$$

by Lemma 2.4, $\Phi(x^+, s^+, (x^+)^T s^+/n) = \eta(\tau_0, n) < \eta(\tau, n)$ for some $\tau_0 < \tau$. By the definition of $\theta(\cdot, \cdot)$, we have $\mu^+ = \theta(\tau_0, n)(x^+)^T s^+/n$. Since $\theta(\cdot, n)$ is a decreasing function of τ, we have $\theta(\tau_0, n) > \theta(\tau, n)$ and $\mu^+ \geq \theta(\tau, n)(x^+)^T s^+/n$. Moreover, one has

$$\frac{(x^+)^T s^+}{\mu^+} = \frac{n}{\theta(\tau_0, n)} < \frac{n}{\theta(\tau, n)},$$

which implies (3.16). This completes the proof of the lemma. ∎

Now we can progress to estimate the increase of the proximity function in the predictor step. To release the notational load, we drop the subscript + used in our previous discussion and simply denote the present iterate by (x, s) and correspondingly the scaled vector by v. We emphasize here again that the scaled vector v satisfies all the inequalities (3.16), (3.17), and (3.18). We start by estimating $v_{\min}$ for the scaled vector v. Since v satisfies all the inequalities from (3.16)–(3.18), by following an analogue chain of reasoning as in the proof of the inequality (3.7) in Lemma 3.2, we obtain

$$v_{\min} \geq \exp^{-\tau}. \tag{3.21}$$

Now we can state the main result of this section.

THEOREM 3.7.

Let (d_x, d_s) be the solution of system (3.4) *with the vector v satisfying properties* (3.16), (3.17), *and* (3.18). *If $n \geq 8$ and $\tau \geq 2$, then the step size*

$$\alpha = \frac{\exp^{1-2\tau}}{15\sqrt{n}\log n}$$

is strictly feasible. Moreover, the new iterate updated with this step size is still in the neighborhood defined by (2.3).

Proof. We start our discussion on the value of the function $\Psi(v(\alpha))$ for a strictly feasible step size α. Note that in the predictor step, all the inequalities (3.16), (3.17), and (3.18) hold. Therefore, we have

$$\max\{\|d_x\|, \|d_s\|\} \leq \|d_x + d_s\| = \|v\| \leq \sqrt{\frac{n}{\theta(\tau, n)}} \leq \exp^{\tau-1}\sqrt{n}, \tag{3.22}$$

where the last inequality follows from (2.11). It follows immediately that

$$\alpha_{\max} \geq v_{\min}\sqrt{\frac{\theta(\tau, n)}{n}} \geq \frac{\exp^{1-\tau} v_{\min}}{\sqrt{n}}.$$

By following a similar reasoning chain as in our discussion in the corrector step, we can show that

$$\begin{aligned}
&\Psi(v(\alpha)) - \Psi(v(0)) \\
&\quad \leq -\alpha\|v\|^2 + \frac{1}{2\log n}\sum_{i=1}^{n}((v + \alpha d_x)_i^{-\log n} + (v + \alpha d_s)_i^{-\log n} - 2v_i^{-\log n}).
\end{aligned}$$

Let

$$h(\alpha) := -\alpha \|v\|^2 + \frac{1}{2\log n}\sum_{i=1}^{n}((v+\alpha d_x)_i^{-\log n} + (v+\alpha d_s)_i^{-\log n} - 2v_i^{-\log n}).$$

It is easy to verify that $h(\alpha)$ is a convex function of α. Furthermore, by simple calculus one gets

$$h(0) = 0 \quad \text{and} \quad h'(0) = \sum_{i=1}^{n} {v_i}^{-\log n} - \|v\|^2.$$

From (3.17), we can claim that $h'(0) \le 0$. Moreover, for any $0 < \alpha \le \alpha_{\max}$, one has

$$\begin{aligned} h''(\alpha) &= (1+\log n)\sum_{i=1}^{n}((d_x^i)^2(v+\alpha d_x)_i^{-2-\log n} + (d_s^i)^2(v+\alpha d_s)_i{}^{-2-\log n}) \\ &\le (1+\log n)\sum_{i=1}^{n}(v_i - \alpha\|v\|)^{-2-\log n}\big((d_x^i)^2 + (d_s^i)^2\big) \\ &\le (1+\log n)\|v\|^2(v_{\min} - \alpha\|v\|)^{-2-\log n}. \end{aligned}$$

It follows that for any $\alpha \in (0, \alpha_{\max}]$,

$$\begin{aligned} \Psi(v(\alpha)) &\le \Psi(v) + \int_0^\alpha \int_0^\zeta (1+\log n)\|v\|^2(v_{\min} - t\|v\|)^{-2-\log n} dt d\zeta \\ &< \Psi(v) + \int_0^\alpha \|v\|\big((v_{\min} - \zeta\|v\|)^{-1-\log n}\big)d\zeta \\ &= \Psi(v) + \frac{(v_{\min} - \alpha\|v\|)^{-\log n} - v_{\min}^{-\log n}}{\log n}. \end{aligned}$$

By using (3.18) we obtain

$$\Psi(v) \le \eta_1(\tau, n) - \frac{5}{66\log n} v_{\min}^{-\log n}.$$

Therefore, for any step size α satisfying

$$(v_{\min} - \alpha\|v\|)^{-\log n} \le \frac{5v_{\min}^{-\log n}}{66} + v_{\min}^{-\log n}, \tag{3.23}$$

one has

$$\Psi(v(\alpha)) \le \eta_1(\tau, n),$$

which, by Lemma 2.4, further implies that

$$\Phi(x(\alpha), s(\alpha), \mu(x,s)(\alpha)) \le \eta(\tau, n).$$

Thus for any step size α satisfying (3.23), the new iterate $(x(\alpha), s(\alpha))$ is still in the neighborhood defined by (2.3). Now we are going to estimate the maximal step size α^* that satisfies (3.23), which is the unique solution of the following equation:

$$(1 - \alpha v_{\min}^{-1}\|v\|)^{-\log n} = \frac{71}{66}. \tag{3.24}$$

The above relation implies that

$$-\log\left(1-\alpha^* v_{\min}^{-1}\|v\|\right) = \frac{\log\frac{71}{66}}{\log n} \geq \frac{1}{14\log n}. \tag{3.25}$$

Using Lemma 3.3 and the fact that $\log n \geq 1$, we have

$$\alpha^* v_{\min}^{-1}\|v\| \geq \frac{1}{1+14\log n} > \frac{1}{15\log n}.$$

It follows immediately that

$$\alpha^* \geq \frac{v_{\min}}{15\|v\|\log n} \geq \frac{\exp^{1-\tau} v_{\min}}{15\sqrt{n}\log n} \geq \frac{\exp^{1-2\tau}}{15\sqrt{n}\log n},$$

where the second inequality follows from (3.22) and the last inequality from (3.21). The proof of the theorem is finished. □

We close this section by giving an upper bound for the number of iterations of Algorithm 3.1. As indicated in Proposition 3.1, the duality gap will not increase in the corrector step. Theorem 3.7 shows that the duality gap will decrease at least at a rate of

$$1-\frac{\exp^{1-2\tau}}{15\sqrt{n}\log n}$$

in each predictor step. As a direct consequence of these two results, we have the following theorem.

THEOREM 3.8. *If $n \geq 8$ and $\tau \geq 2$, then after at most*

$$\left\lceil 15\exp^{2\tau-1}\sqrt{n}\log n\log\frac{(x^0)^T s^0}{\epsilon}\right\rceil$$

iterations, Algorithm 3.1 *will find an approximate solution satisfying $x^T s \leq \epsilon$.*

We note that, although Algorithm 3.1 has an $O(\sqrt{n}\log n\log\frac{(x^0)^T s^0}{\varepsilon})$ complexity, the upper bound of the number of total iterations given by Theorem 3.8 is exponential in τ.

4. Local superlinear convergence. In this section we prove the local superlinear convergence of the algorithm. Let (x_-, s_-) denote the primal-dual pair before the corrector step and (x, s) the new pair after a corrector step. Recall that for the predictor direction, Ye and Anstreicher [28] have proved the following relation for monotone complementarity problems:

$$|\Delta x_i \Delta s_i| = O((\mu(x,s))^2), \quad i = 1,\ldots,n, \tag{4.1}$$

when the product $x^T s$ (or $\mu(x,s)$) is sufficiently small. Since an LO problem can be cast as a special class of monotone linear complementarity problems, the above relation remains valid for LO if $\mu(x,s)$ is small enough. To simplify the analysis, we

ignore the hidden constant in the big-O relation. Now let us recall the choice of the parameter μ in the predictor step in our algorithm. We then have

$$|d_x^i d_s^i| = \frac{|\Delta x_i \Delta s_i|}{\mu} \leq \frac{|\Delta x_i \Delta s_i|}{\theta(\tau, n)\mu(x,s)} = O(\mu(x,s)), \tag{4.2}$$

where the inequality follows from Lemma 3.6. We next give a technical result about the step size used in the predictor step.

LEMMA 4.1. *Let $n \geq 8$, $\tau \geq 2$, $r \in (0,1)$, and (x,s) be an iterate in Algorithm* 3.1. *If the present duality gap $x^T s$ is sufficiently small so that the relation* (4.1) *holds, then the step size α used in the predictor step satisfies $\alpha \geq 1 - O((\mu(x,s))^r)$.*

Proof. We start with an estimation of the maximal feasible step size. As we mentioned in section 3, the new iterate is strictly feasible if and only if both $v + \alpha d_x$ and $v + \alpha d_s$ are strictly feasible. Further, the maximal feasible step size $\alpha_{\max}$ satisfies the condition

$$(v + \alpha d_x)(v + \alpha d_s) = v^2 - \alpha v^2 + \alpha^2 d_x d_s \geq 0, \quad \alpha \in [0, \alpha_{\max}],$$

or

$$e - \alpha + \alpha^2 v^{-2} d_x d_s \geq 0 \quad \forall \alpha \in [0, \alpha_{\max}].$$

Since $v_{\min} \geq \exp^{-\tau}$ and $|d_x d_s| = O(\mu(x,s))$, we can conclude that $\alpha_{\max} \geq 1 - O(\mu(x,s))$.

Now we are going to show that for any fixed $0 < r < 1$, the step size α used in the predictor step also satisfies the relation

$$\alpha \geq 1 - O((\mu(x,s))^r). \tag{4.3}$$

To prove (4.3), we notice that for any strictly feasible step size $\alpha \in (0,1]$, from the choice of μ and the corresponding scaled vector v we obtain

$$\begin{aligned}
&\Phi(x(\alpha), s(\alpha), (1-\alpha)\mu) - \Phi(x(0), s(0), \mu) \\
&\quad \leq \frac{1}{\log n} \sum_{i=1}^{n} \left(\left(\frac{v_i(\alpha)}{\sqrt{1-\alpha}} \right)^{-\log n} - {v_i}^{-\log n} \right) \\
&\quad = \sum_{i=1}^{n} \frac{v_i^{-\log n}}{\log n} \left(\left(\frac{v_i(\alpha)}{\sqrt{1-\alpha} v_i} \right)^{-\log n} - 1 \right) \\
&\quad = \sum_{i=1}^{n} \frac{v_i^{-\log n}}{\log n} \left(\left(\frac{v_i^2(\alpha)}{(1-\alpha) v_i^2} \right)^{-\frac{1}{2}\log n} - 1 \right) \\
&\quad = \sum_{i=1}^{n} \frac{v_i^{-\log n}}{\log n} \left(\left(1 + \frac{\alpha^2}{1-\alpha} v_i^{-2} d_x^i d_s^i \right)^{-\frac{1}{2}\log n} - 1 \right).
\end{aligned}$$

Let $\mathcal{I}_-$ be the index set

$$\mathcal{I}_- = \{ i \in \{1, \ldots, n\} : d_x^i d_s^i < 0 \}.$$

It follows immediately that

$$\begin{aligned}
&\Phi(x(\alpha), s(\alpha), (1-\alpha)\mu) - \Phi(x(0), s(0), \mu) \\
&\quad \le \sum_{i \in \mathcal{I}_-} \frac{v_i^{-\log n}}{\log n} \left(\left(1 + \frac{\alpha^2}{1-\alpha} v_i^{-2} d_x^i d_s^i \right)^{-\frac{1}{2}\log n} - 1 \right) \\
&\quad \le \frac{v_{\min}^{-\log n}}{\log n} \sum_{i \in \mathcal{I}_-} \left(\left(1 + \frac{\alpha^2}{1-\alpha} v_i^{-2} d_x^i d_s^i \right)^{-\frac{1}{2}\log n} - 1 \right) \\
&\quad \le \frac{v_{\min}^{-\log n}}{\log n} \left(\left(1 + \sum_{i \in \mathcal{I}_-} \frac{\alpha^2}{1-\alpha} v_i^{-2} d_x^i d_s^i \right)^{-\frac{1}{2}\log n} - 1 \right) \\
&\quad \le \frac{v_{\min}^{-\log n}}{\log n} \left(\left(1 - \frac{\alpha^2}{1-\alpha} O(\mu(x,s)) \right)^{-\frac{1}{2}\log n} - 1 \right).
\end{aligned}$$

This inequality indicates that, for any fixed $r \in (0,1)$ and sufficiently small $\mu(x,s)$, there exists a step size $\alpha \ge 1 - O((\mu(x,s))^r)$ such that

$$\Phi(x(\alpha), s(\alpha), (1-\alpha)\mu) - \Phi(x(0), s(0), \mu) \le \frac{5 v_{\min}^{-\log n}}{66 \log n},$$

as stated in Theorem 3.5. This completes the proof of the lemma. ∎

It is straightforward to verify that if the step size used in the predictor step satisfies $\alpha \ge 1 - O((\mu(x,s))^r)$, then after the predictor step we have

$$\mu_{gap}^{+} = (1-\alpha)\mu(x,s) = O((\mu(x,s))^{1+r}).$$

Now we can state the main result of this section.

THEOREM 4.2. *Let $n \ge 8$, $\tau \ge 2$, and (x^k, s^k) be generated by Algorithm* 3.1. *The algorithm is superlinearly convergent in the sense that $\mu_{gap}^{k+1} = O((\mu_{gap}^k)^{1+r})$ for any fixed $r \in (0,1)$ and every accumulation point of the sequence (x^k, s^k) is a strictly complementary solution of the problem.*

Proof. The superlinear convergence of the algorithm follows from Lemma 4.1. The convergence properties of every accumulation point of the iterates can be proved by using the properties of the central path, which converges to a strictly complementarity solution of the underlying problem. The details of the proof are omitted here and we refer the readers to [27] for analogous discussions. ∎

5. Conclusions. A new predictor-corrector algorithm working in a large neighborhood of the central path is proposed for linear optimization. This algorithm maintains the simple structure of the MTY method, that is, each iteration consists of one predictor step and only one corrector step, and uses only first-order information of the problem. We prove that the algorithm retains local superlinear convergence and has an $O(\sqrt{n}\log n \log \frac{(x^0)^T s^0}{\varepsilon})$ iteration bound. This complexity result improves the so-far best known complexity $O(n \log \frac{(x^0)^T s^0}{\varepsilon})$ for first-order predictor-corrector algorithms working in large neighborhoods [15].

We mention that it is possible to extend the results in this paper to other cases, e.g., predictor-corrector methods for complementarity problems. The details for such an extension are left to the interested reader.

Acknowledgments. The authors would like to thank two anonymous referees for their useful comments. They are grateful for Dr. C. Roos for his invaluable suggestions that led to substantial improvement in the presentation of the paper.

REFERENCES

[1] E. D. Andersen, J. Gondzio, Cs. Mészáros, and X. Xu, *Implementation of interior point methods for large scale linear programs*, in Interior Point Methods of Mathematical Programming, T. Terlaky, ed., Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996, pp. 189–252.

[2] K. M. Anstreicher and R. A. Bosch, *A new infinity-norm path following algorithm for linear programming*, SIAM J. Optim., 5 (1995), pp. 236–246.

[3] C. C. Gonzaga, *Complexity of predictor-corrector algorithms for LCP based on a large neighborhood of the central path*, SIAM J. Optim., 10 (1999), pp. 183–194.

[4] C. C. Gonzaga and R. A. Tapia, *On the convergence of the Mizuno–Todd–Ye algorithm to the analytical center of the solution set*, SIAM J. Optim., 7 (1997), pp. 47–65.

[5] P. F. Huang and Y. Ye, *An asymptotical $O(\sqrt{n}L)$-iteration path-following linear programming algorithm that uses wide neighborhoods*, SIAM J. Optim., 6 (1996), pp. 570–586.

[6] B. Jansen, C. Roos, T. Terlaky, and J. P. Vial, *Primal-dual path-following algorithms for linear programming*, Ann. Oper. Res., 21 (1996), pp. 197–232.

[7] B. Jansen, C. Roos, T. Terlaky, and Y. Ye, *Improved complexity using higher-order corrections for primal-dual Dikin affine scaling*, Math. Programming, Series B, 76 (1997), pp. 117–130.

[8] N. K. Karmarkar, *A new polynomial-time algorithm for linear programming*, Combinatorica, 4 (1984), pp. 373–395.

[9] M. Kojima, N. Megiddo, and S. Mizuno, *Theoretical convergence of large-step primal-dual interior point algorithms for linear programming*, Math. Programming, Series A, 59 (1993), pp. 1–21.

[10] Z.-Q. Luo, J. F. Sturm, and S. Zhang, *Superlinear convergence of a symmetric primal-dual path-following algorithm for semidefinite programming*, SIAM J. Optim., 8 (1998), pp. 59–81.

[11] S. Mizuno, M. J. Todd, and Y. Ye, *On adaptive-step primal-dual interior-point algorithms for linear programming*, Math. Oper. Res., 18 (1993), pp. 964–981.

[12] R. C. Monterio, I. Adler, and M. G. C. Resende, *A polynomial-time primal-dual affine scaling algorithm for linear and convex quadratic programming and its power series extension*, Math. Oper. Res., 15 (1990), pp. 191–214.

[13] J. M. Peng, C. Roos, and T. Terlaky, *Self-regular proximities and new search directions for linear and semidefinite optimization*, Math. Programming, Series A, 93 (2002), pp. 129–171.

[14] J. M. Peng and T. Terlaky, *A dynamic large-update primal-dual interior-point method for linear optimization*, Optim. Methods Softw., 17 (2002), pp. 1077–1104.

[15] F. A. Potra, *The Mizuno-Todd-Ye algorithm in a larger neighborhood of the central path*, European J. Oper. Res., 143 (2002), pp. 257–267.

[16] F. A. Potra and R. Sheng, *A large-step infeasible-interior-point method for the P_*-matrix LCP*, SIAM J. Optim., 7 (1997), pp. 318–335.

[17] F. A. Potra, *A Superlinear Convergent Predictor-Corrector Method for Degenerate LCP in a Wide Neighborhood of the Central Path with $O(\sqrt{n}L)$-Iteration Complexity*, Technical report, Department of Mathematics and Statistics, University of Maryland, College Park, MD, 2001.

[18] J. Renegar, *A Mathematical View of Interior-Point Methods in Convex Optimization*, MPS/SIAM Ser. Optim. 3, SIAM, Philadelphia, 2001.

[19] C. Roos, T. Terlaky, and J. Ph. Vial, *Theory and Algorithms for Linear Optimization: An Interior Approach*, John Wiley & Sons, Chichester, UK, 1997.

[20] J. Stoer and M. Wechs, *The complexity of high-order predictor-corrector methods for solving sufficient linear complementarity problems*, Optim. Methods Softw., 10 (1998), pp. 393–417.

[21] J. Stoer, W. Wechs, and S. Mizuno, *High order infeasible-interior-point methods for solving sufficient linear complementarity problems*, Math. Oper. Res., 23 (1998), pp. 832–862.

[22] J. Sturm, *Theory and algorithms of semidefinite programming*, in High Performance Optimization, H. Frenk, C. Roos, T. Terlaky, and S. Zhang, eds., Kluwer Academic Publishers, Dordrecht, The Netherlands, 2000, pp. 1–194.

[23] J. Sturm and S. Zhang, *On a wide region of centers and primal-dual interior-point algorithms*

for linear programming, Math. Oper. Res., 22 (1997), pp. 408–431.

[24] J. Sun and G. Zhao, *Global linear and local quadratic convergence of a long-step adaptive-mode interior point method for some monotone variational inequality problems*, SIAM J. Optim., 8 (1998), pp. 123–139.

[25] L. Tüncel, *On the convergence of primal-dual interior-point methods with wide neighborhoods*, Comput. Optim. Appl., 4 (1995), pp. 139–158.

[26] S. J. Wright, *Primal-Dual Interior-Point Methods*, SIAM, Philadelphia, 1997.

[27] Y. Ye, *Interior-Point Algorithms, Theory and Analysis*, John Wiley and Sons, Chichester, UK, 1997.

[28] Y. Ye and K. Anstreicher, *On quadratic and $O(\sqrt{n}L)$ convergence of a predictor-corrector algorithm for LCP*, Math. Programming, Series A, 62 (1993), pp. 537–551.

[29] Y. Ye, O. Güler, R. A. Tapia, and Y. Zhang, *A quadratically convergent $O(\sqrt{n}L$-iteration algorithm for linear programming*, Math. Programming, Series A, 59 (1993), pp. 151–162.

[30] Y. Zhang and D. Zhang, *On polynomiality of the Mehrotra-type predictor-corrector interior point algorithms*, Math. Programming, Series A, 68 (1995), pp. 303–318.

[31] G. Zhao, *Interior-point algorithms for linear complementarity problems based on large neighborhoods of the central path*, SIAM J. Optim., 8 (1998), pp. 397–413.